

ORIGINAL RESEARCH

Identification and Validation of Prognostic Risk Model for Female-Specific Lung Adenocarcinoma

Siyang Feng, MD; Yunhui Guo, MS; Jianxue Zhai, MS; Siyu Xie, MS; Wei Yang, PhD

ABSTRACT

Background • Lung adenocarcinoma (LUAD) is a major pathological subtype of non-small cell lung cancer and occurs more commonly in females than other lung cancer subtypes. Studying female-specific oncogenes in LUAD may provide personalized medicine approaches for females with LUAD.

Objective • We aimed to identify the possible female-specific oncogenes of LUAD and understand their potential impact on treatment strategies for specific cancer subgroups.

Methods • The gene expression profiles of LUAD were downloaded from The Cancer Genome Atlas (TCGA) database and the GSE72094 dataset. TCGA database is currently the largest database of cancer genetic information. Female-specific differentially expressed genes (DEGs) were identified by R programming software. Functional annotation of DEGs was conducted based on KEGG pathway enrichment analysis. Univariate and multivariate Cox proportion analyses were applied to construct a prognostic risk score model with the DEGs. Kaplan-Meier and ROC curves were plotted to validate the predictive effect of the prognostic DEGs signature. Gene set enrichment analysis (GSEA) was applied to identify the potential pathways in the high-risk groups in female LUAD. Finally, the immunohistochemical staining (IHC) was conducted to verify the expression of CABLES1 in human LUAD samples.

Results • We constructed a prognostic signature that includes 12 female-

specific DEGs ($P < .05$). Among them, ABHD6, CABLES1, CXCL5, DNAJB4, EFNB2, HLX, MEOX2, MTMR10, PPFIBP1, and RERG were down-regulated in LUAD, while MFSD6L and SOX9 were up-regulated in LUAD ($P < .0001$). The Kaplan-Meier, and receiver operator characteristic (ROC) curves revealed efficient and stable prediction of the prognostic signature in the female LUAD patients. It was showed the risk score model has a good predictive effect on the prognosis of female LUAD patients but is not effective for male patients ($P < .0001$). The ROC curve showed that the areas under the curve (AUC) of first-, third- and fifth-year survival were 0.70, 0.69, and 0.79, respectively, which indicated good sensitivity and specificity of the 12-gene risk score algorithm in predicting the prognosis of female LUAD. GSEA revealed that the high-risk group was significantly enriched in the EMT, E2F targets, Myc targets, G2/M checkpoint, glycolysis, hypoxia, and mTORC1 signaling pathways ($P < .05$). Immunohistochemical staining showed lower CABLES1 expression was associated with higher pTNM stage in female LUAD but not in male LUAD ($P < .05$).

Conclusion • Our study constructed and verified a prognostic signature based on 12 female-specific DEGs of LUAD, which could improve the understanding of sex-related risk factors involved in LUAD carcinogenesis and progression, and may provide personalized treatment strategies for female LUAD patients. (*Altern Ther Health Med*. [E-pub ahead of print.]

Siyang Feng, MD; Yunhui Guo, MS; Siyu Xie, MS; Wei Yang, PhD, Department of Pathology; Guangdong Province Key Laboratory of Molecular Tumor Pathology; School of Basic Medical Sciences; Southern Medical University; Guangzhou; China. **Yunhui Guo, MS; Wei Yang, PhD**, Research Department of Medical Sciences; Guangdong Provincial People's Hospital; Guangdong Academy of Medical Sciences; Southern Medical University, Guangzhou; China. **Siyang Feng, MD; Jianxue Zhai, MS**, Department of Thoracic Surgery; Nanfang Hospital; Southern Medical University; Guangzhou; China.

Corresponding author: Wei Yang, PhD

E-mail: yanglabgz@163.com

INTRODUCTION

Lung cancer stands as the primary contributor to cancer-related fatalities globally.¹ Smoking is intricately linked to the prevalence of lung squamous cell carcinoma (LUSC).^{2,3} However, non-tobacco-related factors, such as sex and hormones, may wield significant influence over the onset and

progression of lung adenocarcinoma (LUAD).⁴ Notably, a higher incidence of LUAD is observed in females, irrespective of smoking habits. Moreover, LUAD is more prevalent among premenopausal women compared to postmenopausal women (58% vs. 47%).⁵ Insights from a prospective cohort of 36,588 peri- and postmenopausal women suggest that hormone replacement therapy (HRT), comprising estrogen plus progestin, heightens the risk of lung cancer in a duration-dependent manner, presenting an approximately 50% elevated risk after 10 years or more.⁶ Furthermore, a randomized controlled trial involving 16,608 women underscores the elevated mortality rate among lung cancer patients treated with combined HRT (estrogen plus progestin) compared to the placebo group.⁷ Recent systematic analysis from 18 studies confirms a significant increase in the risk of LUAD associated with HRT in nonsmoking women, with an attributable risk of 76%.⁸ Additionally, recent observations reveal crosstalk between estrogen receptor pathways and epidermal growth factor receptor (EGFR) in LUAD.⁹ EGFR, a receptor tyrosine kinase, constitutes a pivotal molecular target for LUAD. Estrogen signaling induces epidermal

growth factor (EGF) production, thereby activating EGFR signaling.^{10,11} Despite these advancements, gender-related carcinogenic factors (including sex hormones and associated receptors and their related molecular pathways) and effective medical interventions for LUAD remain insufficient. Sex differences in LUAD biology are based on intrinsic genetic differences, as well as overlapping epigenetic changes and the influence of sex hormones.¹² A better understanding of the mechanisms of gender differences in tumor immunotherapy could allow gender factors to be considered in tumor drug development, leading to more targeted treatment strategies.

Understanding the landscape of gene expression alterations in lung adenocarcinoma (LUAD) is paramount for advancing prognostic models in oncology. Differentially expressed genes (DEGs) play a pivotal role in delineating molecular signatures that correlate with disease progression, treatment response, and patient outcomes. Identifying genes whose expression levels significantly deviate between tumor and normal tissues can unravel key molecular pathways implicated in LUAD pathogenesis. Integrating DEGs into predictive models enhances the accuracy of patient stratification and provides insights into the underlying biological mechanisms driving disease aggressiveness. As reported previously, sex-biased competitive endogenous (ce) RNA networks reveal that OSCAR can promote proliferation and migration of LUAD in women.¹³ Consequently, leveraging DEGs in prognostic modeling holds promise for refining risk assessment, guiding personalized therapeutic strategies, and ultimately improving the clinical management of LUAD patients. However, few articles have been reported about possible female-specific oncogenes of LUAD.

With the development of the Cancer Genome Atlas (TCGA) database based on high-throughput sequencing, clinicians and researchers have gained a growing understanding of the pathogenesis of various cancers.¹⁴ TCGA data are important for guiding the prevention, diagnosis and treatment of cancers.¹⁵ The TCGA database contains more than 20,000 normally matched samples of primary cancer and 33 cancer types, including gene expression data, DNA methylation data, and standardized clinical data, which are important for clinicians and researchers to understand the mechanisms of related cancers and to discover potential prognostic biomarkers.¹⁶

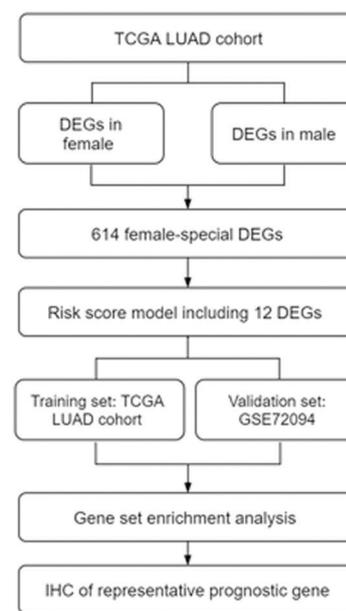
In this study, we aimed to construct an effective prognostic risk model based on DEGs for female LUAD using bioinformatic methods, and validate the predictive effect of the prognostic DEGs signature and identify the potential pathways in the high-risk group of female LUAD patients, which might provide new insight into the molecular pathogenesis of LUAD.

MATERIALS AND METHODS

Data acquisition and preparation

The workflow of the study is displayed in Figure 1. The RNA sequencing data of LUAD and LUSC samples in The Cancer Genome Atlas (TCGA) cohort were acquired via the

Figure 1. Workflow chart of prognostic-risk model of female LUAD



UCSC Xena browser (<https://xenabrowser.net/>), which functioned as a training set. A total of 519 LUAD and 500 LUSC patients with available sexual information were included in the analysis. TCGA is a landmark cancer genomics program that has generated comprehensive, multi-dimensional maps of key genomic changes in various types and subtypes of cancer. Another dataset of LUAD (GSE72094) was downloaded from the Gene Expression Omnibus (GEO) database (<http://ncbi.nlm.nih.gov/geo/>) and served as a validation set. GEO is a public repository that archives and freely distributes high-throughput gene expression and molecular abundance data, including microarray and next-generation sequencing experiments. The gene expression data were annotated based on R software (R-4.0.4-win). The collected clinicopathological data included age, history of smoking, TNM stage, survival status, and survival duration in days. Our research excluded any samples that had missing or insufficient data on age, grade, stage, survival status, and survival duration.

Functional enrichment of KEGG based on female-specific differentially expressed genes

The RNA-seq data of TCGA-LUAD primary solid tumor and solid tissue normal samples were screened to clear untrusted data. The limma analyses were applied to identify the differential expressed genes (DEGs) in female LUAD and male LUAD patients. Limma is a linear model-based differential analysis method that was originally developed for gene-chip data, but has since been applied to RNA sequencing data as well. It uses the weighted least squares method to estimate differences in gene expression and corrects for multiple test problems through Bayesian methods. limma generally assumes a normal distribution of gene expression data, performs well when processing large-scale data, and is

suitable for high-throughput data analysis, such as chip and large-scale RNA sequencing data, with good control of false positive rates.¹⁷ The DEGs with $|\log_2$ fold change (FC)| > 1 and an adjusted $P < .05$ were considered significant and visualized using volcano plots. In addition, the intersection between female DEGs and male DEGs was shown in a Venn diagram. To further predict the potential functional enrichment of female-specific DEGs, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using the DAVID database (<https://david.ncicrf.gov/>).^{18,19}

Construction and validation of the predictive risk score model in female LUAD patients

A univariate Cox regression analysis was performed to evaluate the overall survival (OS)-associated female-specific DEGs in the training set. A total of 269 female LUAD samples with matched survival information were included for survival analysis. DEGs with $P < .05$ were further studied. The univariate Cox regression analysis result showed that 35 female-specific DEGs were significantly associated with OS. A backward stepwise (according to the Akaike information criterion [AIC]) multivariate Cox proportion analysis was conducted to obtain the optimal prognostic genes for the model. The following formula was applied to calculate the risk score of each patient, which combined regression coefficients and expression values of each gene: risk score = (index gene 1 × expression of gene 1) + (index gene 2 × expression of gene 2) + + (index gene 12 × expression of gene 12). All female and male LUAD patients in the TCGA cohort were divided into two subgroups (high-risk group and low-risk group) according to the median risk score. A heatmap of 12 DEGs in the risk score model was created with the pheatmap package. Kaplan–Meier curves were used to demonstrate the survival function of the two groups. The risk model's accuracy for predicting patients' first-, third- and fifth-year survival proportions was estimated using the receiver operating characteristic (ROC) curve. Subsequently, the prognostic model was validated in the GSE72094 dataset. Batch effects between TCGA and GEO data were removed by creating precise models using the "sva" package in R.

Gene set enrichment analysis

To reveal the related pathways and molecular mechanisms of the the high-risk gene set of female LUAD patients, gene set enrichment analysis (GSEA)²⁰ was performed using the GSEA 4.2.3 software (<https://www.gsea-msigdb.org/gsea/index.jsp>). The online gene set named "h.all.v7.5.1 symbols.gmt [Hallmarks]" was used in the GSEA calculation. Gene sets with $P < .05$ and a false discovery rate (FDR) < 0.25 after 1,000 permutations were considered significantly enriched. An FDR < 0.25 indicates that the result is likely to be valid 3 out of 4 times, which is reasonable in the setting of exploratory discovery where one is interested in finding candidate hypothesis to be further validated as a results of future research.

Immunohistochemistry staining for the expression of CABLES1 in LUAD

Since high expression of CABLES1 was correlated with longer OS of male and female LUAD patient, we further evaluated the expression of CABLES1 in LUAD. A total of 151 patients with LUAD from Nanfang Hospital, Southern Medical University (Guangzhou, China) were included for immunohistochemistry analysis. The study followed the ethical guidelines of the Declaration of Helsinki and was approved by the Ethics Committee of Nanfang Hospital. All patients signed informed consent. Paraffin-embedded LUAD tissues were deparaffinized and rehydrated. The antigens were retrieved with 10 mmol/L sodium citrate (pH=6.0). After incubation with 3% H₂O₂ and blocking with goat serum, the slices were incubated with primary antibody (rabbit polyclonal antibody to CABLES1, 1:100, AP60124, Abcepta) at 4°C overnight. The next day, the slides were incubated with an HRP-conjugated goat anti-rabbit IgG antibody (1:2000, ab205718, Abcam). The color reaction was performed with 3-amino-9-ethylcarbazole (AEC) (A2010, Solarbio, China). Finally, the hematoxylin was used to counterstain the sections. The slides were photographed using an Olympus BX53 microscope. Immunohistochemical staining results were scored in accordance with immunoreactive score (IRS) standards proposed by Remmele and Stegner²¹, in which IRS = SI (staining intensity) × PP (percentage of positive cells). Negative PP was defined as 0, 0–10%, 10%–50%, 51%–80%, and >80% PP was defined as 1, 2, 3, and 4, respectively. Negative, mild, moderate, and strongly positive SI was defined as 0, 1, 2, and 3, respectively.

Statistical analysis

All our data processing and picture drawing was carried out by using R software. GraphPad Prism 9.0 software was used to analyze and visualize the statistical profile. Two-tailed Student's *t* test analyzed the differences of gene expression data with normal distribution between tumor and normal samples. Kaplan–Meier curves were used to describe survival data, and a two-sided log-rank test was used to compare data with abnormal distribution between the two groups. The $P < .05$ was considered significant.

RESULTS

Functional enrichment of female-specific DEGs in LUAD

A total of 519 LUAD and 500 LUSC patients with available sexual information were included in the analysis. There were 279 female (188 smokers and 91 nonsmokers) and 240 male (179 smokers and 61 nonsmokers) patients in the LUAD cohort and 130 female (115 smokers and 15 nonsmokers) and 370 male (316 smokers and 54 nonsmokers) patients in the LUSC cohort. LUSC exhibited a strong male predominance in incidence ($P > .05$) (Figure 2A). There were 275 primary LUAD tissues 33 normal lung tissues from females and 235 primary LUAD tissues, and 25 normal lung tissues from males. A total of 4496 DEGs in female LUAD patients and 7347 DEGs in male LUAD patients were

Figure 2. Screening of female-specific DEGs in LUAD patients from the TCGA database. A. Sex composition of LUAD and LUSC in the TCGA cohort. B. Volcano plot of DEGs between females and males in LUAD. C. Venn diagram between females and males in LUAD. D. Bubble plot of KEGG pathway enrichment analysis of 614 female-specific DEGs.

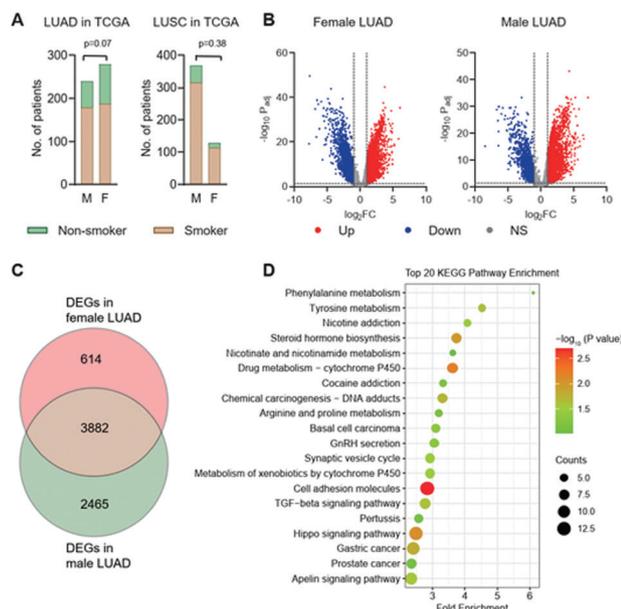
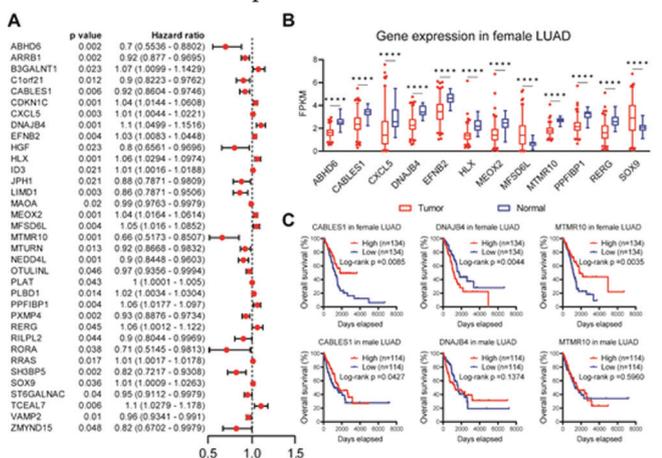


Figure 3. Identification of candidate genes associated with the prognosis of female-specific LUAD patients. A. Univariate Cox regression analysis of female-specific DEGs. B. Box plots of the expression of 12 prognosis-related female-specific DEGs in female LUAD based on multivariate Cox proportion analysis. The gene expression data between the two groups were analyzed by two-tailed Student's *t*-test, *****P* < .0001. C. Kaplan-Meier curves of the relationship between *CABLES1*, *DNAJB4*, *MTMR10* expression and the survival rate of female or male LUAD patients in the TCGA cohort.



identified (Figure 2B) (*P* < .05). Figure 2C showed 3882 intersected DEGs between females and males and 614 female-specific DEGs in LUAD (*P* < .05). In addition, the 614 female-specific DEGs were enriched in the cell adhesion

molecules, the Hippo signaling pathway, and steroid hormone biosynthesis-related signaling pathway based on KEGG analysis (Figure 2D).

Construction of a prognostic risk score model for female LUAD patients

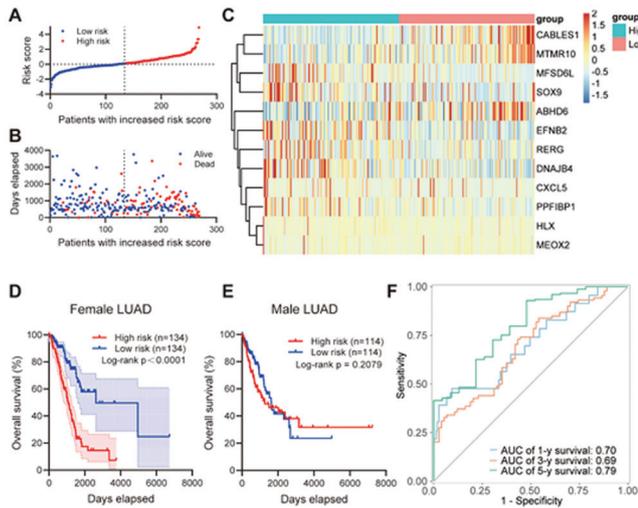
The construction of a prognostic risk score model is beneficial for the early diagnosis and accurate prognostic prediction of female LUAD patients. A total of 269 female LUAD samples with matched survival information were included for survival analysis. The univariate Cox regression analysis result showed that 35 female-specific DEGs were significantly associated with OS (*P* < .05) (Figure 3A). After multivariate Cox proportion analysis, twelve prognosis-related DEGs were screened for the risk score model. The boxplots showed the expression of 12 prognosis-related female-specific DEGs in the TCGA cohort, including *ABHD6*, *CABLES1*, *CXCL5*, *DNAJB4*, *EFNB2*, *HLX*, *MEOX2*, *MFSD6L*, *MTMR10*, *PPFBP1*, *REGG*, and *SOX9*, between LUAD and normal samples (Figure 3B). The risk score of each LUAD patient in TCGA cohort was assessed as follows: risk score = (-0.29819 × *ABHD6* expression) + (-0.06612 × *CABLES1* expression) + (0.009315 × *CXCL5* expression) + (0.073456 × *DNAJB4* expression) + (0.019709 × *EFNB2* expression) + (0.260376 × *HLX* expression) + (-0.13034 × *MEOX2* expression) + (0.060007 × *MFSD6L* expression) + (-0.23938 × *MTMR10* expression) + (0.04584 × *PPFBP1* expression) + (0.152708 × *REGG* expression) + (0.01833 × *SOX9* expression). Lower expression levels of *CABLES1*, *DNAJB4*, *MTMR10* were observed in female LUAD patients with better prognosis (*P* < .05) (Figure 3C).

The female LUAD patients in the TCGA cohort were divided into two groups (high-risk and low-risk groups) based on the median risk score. The distribution of the risk scores and the relationship between risk scores and survival were illustrated in scatterplots (Figure 4A, B), which showed better survival of female patients with LUAD in the low-risk group. The gene expression profiles of the prognostic risk genes between the high-risk group and low-risk group were displayed in the heatmap in Figure 4C. The Kaplan-Meier curve and log-rank test showed a significantly higher OS probability in the low-risk group (*P* < .0001) (Figure 4D), indicating that the risk score model has a good predictive effect on the prognosis of female LUAD patients but is not effective for male patients (Figure 4E). The ROC curve showed that the areas under the curve (AUC) of first-, third- and fifth-year survival were 0.70, 0.69, and 0.79, respectively (Figure 4F), which indicated good sensitivity and specificity of the 12-gene risk score algorithm in predicting the prognosis of female LUAD.

Evaluation of the prognostic risk score model in an independent GEO dataset

To further validate the accuracy and sensitivity of the prognostic risk score model, GSE72094 was used as an external validation dataset. There were 180 female LUAD samples with matched survival data. Consistently, the high-

Figure 4. Evaluation of the prognostic risk score model in the TCGA cohort. A-B. The distribution of risk scores (A) and survival outcomes of the high-risk and low-risk groups (B). C. Heatmap showed the expression of 12 genes between the two groups in female LUAD. D-E. Kaplan-Meier curves showed the survival rate between the two groups in female and male LUAD. F. The ROC curves for predicting the first-, third- and fifth-year OS of female LUAD patients in the TCGA cohort.



risk group suffered a worse survival prognosis, as revealed by scatterplots and Kaplan-Meier curves ($P < .05$) (Figure 5A, B, and D). The gene expression profiles of the prognostic risk genes between the two groups were shown in Figure 5C, which was similar to the result in the TCGA training cohort. The AUCs of first-, third- and fifth-year survival were 0.57, 0.64, and 0.82, respectively (Figure 5E), indicating the model's reasonably acceptable predictive ability. However, due to the retrospective nature of TCGA and GEO data, we need further validate the risk score model in prospective studies.

Exploration of signaling pathways based on the risk score

We conducted GSEA to reveal the significant signaling pathways in the high-risk group based on the prognosis model in female LUAD. The GSEA revealed that epithelial-mesenchymal transition (EMT) ($FDR < 0.0001$, $P < .0001$), E2F targets ($FDR < 0.0001$, $P < .0001$), Myc targets ($FDR < 0.0001$, $P < .0001$), G2/M checkpoint ($FDR < 0.0001$, $P < .0001$), glycolysis ($FDR < 0.0001$, $P < .0001$), hypoxia ($FDR < 0.0001$, $P < .0001$), apical junction and angiogenesis ($FDR = 0.0003$, $P < .0001$) and mTORC1 signaling pathways ($FDR < 0.0001$, $P < .0001$) were significantly enriched in the high-risk group in female LUAD (Figure 6A and Figure 6B). The occurrence of EMT to contribute to poor prognosis in LUAD²². It has been reported that high expression of E2F8 predicts a low 5-year overall survival rate in LUAD patients.²³ MYC overexpression is a novel form of therapeutic resistance to ROS1 tyrosine kinase inhibitors in LUAD.²⁴ G2/M checkpoint, glycolysis, hypoxia, apical junction and angiogenesis and mTORC1 signaling pathways are involved in the progression of LUAD.²⁵⁻²⁹

Figure 5. Validation of the prognostic risk score model in an independent GEO dataset. A-B. The distribution of risk scores (A) and survival outcomes of the high-risk and low-risk groups (B). C. Heatmap showed the expression of 12 genes between the two groups in female LUAD. D. Kaplan-Meier curves showed the survival rate between the two groups in female LUAD. E. The ROC curves for predicting the first-, third- and fifth-year OS of female LUAD patients in the GSE72094 dataset.

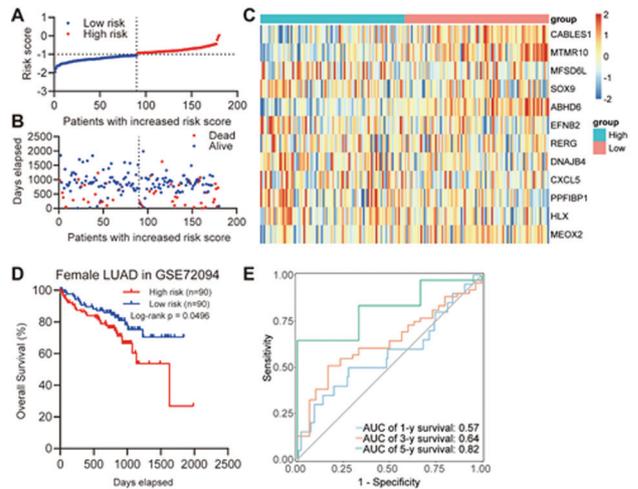
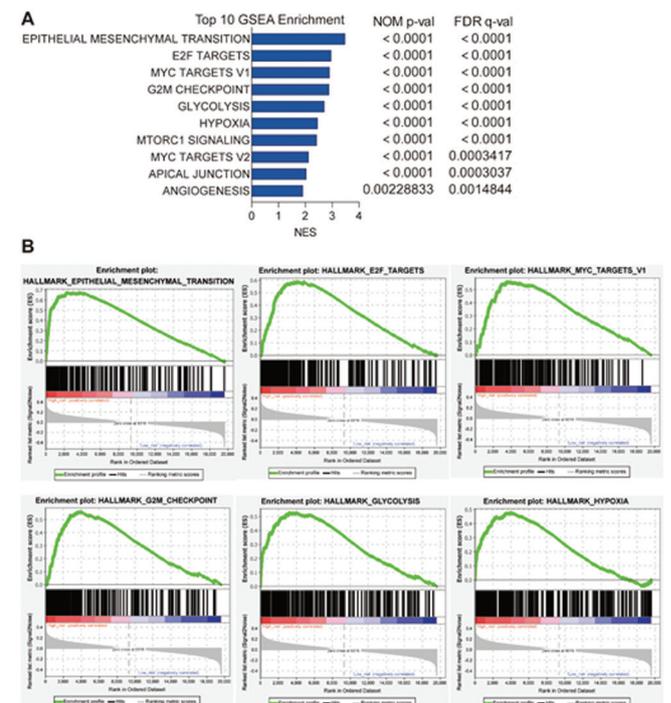


Figure 6. GSEA reveals the enriched signaling pathways in the high-risk group of female LUAD patients in the TCGA cohort. A. Top 10 enriched signaling pathways based on GSEA enrichment score. B. The enrichment plots of the top 6 enriched signaling pathways are based on GSEA enrichment score.

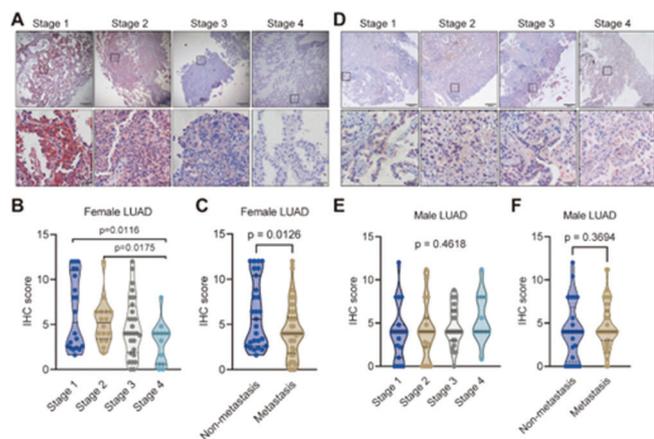


Abbreviations: NOM p-val, nominal P value; FDR q-val, false discovery rate q -value; NES, Normalized enrichment score.

Table 1. Clinical characteristics of LUAD patients in IHC analysis

	Female (n=77) n (%)	Male (n=74) n (%)
Age		
<60	40 (51.95%)	36 (48.65%)
≥60	37 (48.05%)	38 (51.35%)
Smoking status		
Present or former smoker	3 (3.90%)	55 (74.32%)
Non-smoker	74 (96.10%)	19 (25.68%)
pT		
T1	32 (41.56%)	22 (29.73%)
T2	36 (46.75%)	35 (47.30%)
T3	9 (11.69%)	13 (17.57%)
T4	0 (0.00%)	4 (5.41%)
pN		
N0	31 (40.26%)	30 (40.54%)
N1 or N2	40 (51.95%)	40 (54.05%)
Nx	6 (7.79%)	4 (5.41%)
pM		
M0	64 (83.12%)	63 (85.14%)
M1	13 (16.88%)	11 (14.86%)
pTNM stage		
1	20 (25.97%)	19 (25.68%)
2	16 (20.78%)	19 (25.68%)
3	28 (36.36%)	25 (33.78%)
4	13 (16.88%)	11 (14.86%)

Figure 7. Association between the expression of CABLES1 protein and clinical characteristics. A. Immunohistochemical (IHC) staining results of CABLES1 in female LUAD biopsies. B IHC scores of CABLES1 in different pTNM stages of female LUAD (n=77). Scale bar: 200 μm in the upper row and 20 μm in the lower row. C. IHC scores of CABLES1 in primary lesions of female LUAD with or without metastasis. D. IHC staining results of CABLES1 in male LUAD biopsies. E. IHC scores of CABLES1 in different pTNM stages of male LUAD (n=74). Scale bar: 200 μm in the upper row and 20 μm in the lower row. F. IHC score of CABLES1 in primary lesions of male LUAD with or without metastasis.



Expression validation of prognostic gene CABLES1 in 151 LUAD cohorts

To further evaluate the relationship between the expression of CABLES1 protein and clinical characteristics in LUAD, IHC staining was performed on 151 LUAD tumor samples (77 females and 74 males) (Table 1). Decreased CABLES1 was observed in female LUAD with advanced stage (Figure 7A and B). The samples with lymph node metastasis and/or distant metastasis (N stage > 0 and/or M stage > 0) showed decreased CABLES1 expression compared with nonmetastatic tumor samples (Figure 7C). However,

CABLES1 expression did not demonstrate a significant correlation with tumor stage (Figure 7D, E) and metastasis (Figure 7F) in male LUAD ($P > .05$), which indicated the female-specific role of CABLES1 as a tumor suppressor in LUAD and CABLES1 may be a potential biomarker or therapeutic target for female LUAD.

DISCUSSION

LUSC exhibits significantly higher incidence rates in male patients, whereas LUAD demonstrates similar incidence rates between men and women. LUAD tends to metastasize at an early T stage, posing challenges for effective treatment. Therefore, there is a critical need for novel biomarkers and therapeutic targets to delineate specific cancer subgroups and enhance precision treatment strategies. Estrogen has been implicated in LUAD carcinogenesis and progression.⁵⁻⁸ Recent studies have highlighted the interplay between estrogen receptor signaling pathways and EGFR in LUAD.^{9-11,19} However, further investigation is required to elucidate sex-related carcinogenic factors.

Our study utilized LUAD gene expression data from the TCGA and GEO databases to develop a 12-gene signature for constructing a prognostic risk model. The prognostic risk score model, comprising ABHD6, CABLES1, CXCL5, DNAJB4, EFN2, HLX, MEOX2, MFSD6L, MTMR10, PPFIBP1, RERG, and SOX9, demonstrated effectiveness and stability in predicting the prognosis of female LUAD patients, suggesting these biomarkers have the potential to guide personalized treatment decisions such as targeted therapies or hormone-related interventions. Validation with a GEO dataset confirmed the reliability of the model. GSEA revealed that the high-risk group exhibited significant enrichment in pathways associated with epithelial-mesenchymal transition (EMT), E2F targets, Myc targets, G2/M checkpoint, glycolysis, hypoxia, and mTORC1 signaling. CABLES1 was notably associated with a favorable prognosis in female LUAD patients, while DNAJB4 and MTMR10 were associated with a poorer prognosis. Immunohistochemical staining of CABLES1 was conducted in 151 tumor samples, revealing lower expression of CABLES1 in female LUAD patients with higher pTNM stage but not in males.

CABLES1 has been identified as a candidate tumor suppressor gene.³⁰ Loss of CABLES1 expression has been observed in human lung,³¹ colon,³² ovarian,³³ and endometrial cancers.³⁴ Mechanistically, CABLES1 functions as a linker that connects cyclin-dependent kinases (CDKs) with nonreceptor tyrosine kinases (Src, Abl, and Wee1), thereby regulating the activity of CDKs (Cdk2, Cdk3, and Cdk5) through enhancement of their Y15 phosphorylation.^{35,36} Additionally, CABLES1 interacts with p53 and has been reported to enhance p53-induced cell death.^{37,38} Moreover, CABLES1 serves as an antagonist of proteasome subunit alpha type 3 (PSMA3), thereby increasing the stability of the cell cycle regulator p21 protein.^{36,39} Loss of CABLES1 has been demonstrated to promote tumor progression in the ApcMin/+ mouse model and activate the Wnt/β-catenin signaling pathway.⁴⁰ A tissue

microarray study revealed that 45% of NSCLC samples exhibited negative CABLES1 expression among 116 tumor specimen microcores.³¹ Our investigation further revealed downregulation of CABLES1 expression in advanced female LUAD patients. The regulation of CABLES1 and other tumor-associated genes in the risk score model by gender-related carcinogenic factors warrants further investigation.

PPFIBP1 serves as an interacting partner of protein tyrosine phosphatases and has been detected to be fused with ALK in lung carcinoma patients.⁴¹ Moreover, it has been postulated as a potential target of the metastasis-associated protein S100A4.⁴² Multiple chemokines are recognized for their pivotal roles in driving malignant behaviors, spanning cancer initiation, progression, and drug resistance.⁴³ A recent study has identified that CXCL5 was linked to poorer response to immunotherapy for NSCLC.⁴⁴ The Sex-determining region Y (SRY)-related high mobility group box 9 (SOX9) is a crucial transcription factor and has been observed to be overexpressed in various types of tumors. In accordance with our investigation, recent studies have demonstrated elevated levels of SOX9 in cases of lung adenocarcinoma (LUAD) associated with shorter survival times.^{45,46} In addition, SOX9 might facilitate tumor growth by regulating p21 and CDK4, which is potential hallmark of LUAD.⁴⁶ Elevated expression levels of the RAS-like estrogen-regulated growth inhibitor gene (RERG) have been linked to extended survival following surgery among female individuals with malignant pleural mesothelioma rather than males. RERG has been identified as a prognostic biomarker specific to females.⁴⁷ Overexpression of EFNB2 could enhance lung development in a nitrogen-induced congenital diaphragmatic hernia rat model.⁴⁸ In addition, EFNB2 may serve as a marker for esophageal squamous cell carcinoma prognosis.⁴⁹

In contrast to our results, recent research has found that ABHD6 served as the principal monoacylglycerol lipase and functions as an oncogene in NSCLC. In addition, elevated expression of ABHD6 has been associated with advanced TNM stage and poorer prognosis in patients with NSCLC.⁵⁰ The reason for this difference may be that some of the samples in the literature were from lung squamous cell carcinoma, while our samples were all LUAD. DNAJB4 (HLJ1), a member of the HSP40 family, has been previously recognized as a tumor suppressor gene in lung cancer patients. DNAJB4 has been identified as hub genes of LUAD and was downregulated in NSCLC tissues.^{51,52} In addition, elevated HLJ1 expression was correlated with decreased cancer recurrence and extended survival among NSCLC patients.⁵³ MEOX2 could exert genetic and epigenetic modulation on epidermal growth factor receptor inhibitor (EGFR) gene expression, facilitating the progression of lung malignancy.⁵⁴ Aberrantly high methylation of MEOX2 has been identified in lung cancer.⁵⁵ The overexpression of MEOX2 has been shown to promote cisplatin resistance and proliferation of lung tumor cells. MEOX2 has also been correlated with diminished overall survival rates and increased chemoresistance in lung cancer.^{54,56}

No published studies have validated the association between MTMR10, HLX, FSD6L, and LUAD. MTMR10 has been associated with pathological conditions such as interstitial nephritis, karyomegalic encephalopathy, and schizophrenia. MTMR10 has been associated with pathological conditions such as interstitial nephritis, karyomegalic encephalopathy, and schizophrenia.⁵⁷ Downregulation of HLX has been observed in gastric cancer, with its inhibition associated with the downregulation of T-box transcription factors specific to Th1 cells.⁵⁸ Conversely, The, downregulation of HLX in acute myeloid leukemia (AML) has been linked to induction of G0/G1 phase arrest and inhibition of cell proliferation, facilitated by activation of the JAK/STAT signaling pathway.⁵⁹ A high level of MFSD6 was associated with paclitaxel resistance in ovarian cancer.⁶⁰

Our study has some limitations. First, this study was a retrospective study. Second, due to the genetic heterogeneity of LUAD, potential confounding factors were not taken into account in the analysis and were mainly due to the use of LUAD samples from different platforms, which may introduce sampling bias. In addition, gene reporter and gene silencing assays were not conducted in lung cancer cell lines to validate CABLES1 target genes. To address these limitations, future research should prioritize well-designed prospective clinical trials, functional analyses of cell lines or animal models, or studies of the interaction between estrogen signaling and identified DEGs. In addition, we will further discuss the biological or environmental factors of these sex differences, considering aspects such as hormonal influences, genetic predisposition, or lifestyle factors that may influence the different gene expression patterns of male and female LUAD.

Our study might provide guidance for personalized treatment decisions such as targeted therapies or hormone-related interventions for female LUAD patients. The 12-gene prognostic model should be an effective tool for detecting high-risk patients, enabling early treatment to maximally prevent LUAD advancement. While possessing high predictive power, this model also has a small number of genes, reducing the economic burden on patients. Thus, it has great potential for clinical application and transformation. The genes chosen for the model play a very important role in tumor development, suggesting that they can be potential therapeutic targets.

CONCLUSION

In summary, we identified an effective prognostic gene signature consisting of 12 genes that predict the prognosis of female patients with LUAD. This discovery provides novel insights into sex-related risk factors implicated in LUAD carcinogenesis and progression. Our study provides strong evidence for gender-related differences in carcinogenesis and progression of LUAD. Our study may provide an effective molecule-based reference for early detection, risk stratification, and effective treatment strategies for LUAD in women. Our study might provide guidance for personalized treatment decisions such as targeted therapies or hormone-

related interventions for female LUAD patients. Accurate screening of potential benefit populations by effective genetic markers can help improve the success of clinical trials while avoiding unnecessary safety risks for patients who are less likely to benefit. In the near future, we will perform mechanistic studies to understand the roles of the identified genes in LUAD, or clinical trials to test the utility of the gene signature in guiding treatment decisions.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

Siyang Feng and Yunhui Guo contributed equally to this work.

REFERENCES

- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. doi:10.3322/caac.21654. *CA Cancer J Clin*. 2021;71(1):7-33.
- Doll R, Hill AB. Smoking and carcinoma of the lung. doi:10.1136/bmj.2.4682.739. *Br Med J*. 1950;2(4682):739-48.
- Pesch B, Kendzia B, Gustavsson P, et al. Cigarette smoking and lung cancer-related risk estimates for the major histological types from a pooled analysis of case-control studies. doi:10.1002/ijc.27339. *Int J Cancer*. 2012;131(5):1210-9.
- Hsu LH, Chu NM, Kao SH. Estrogen, Estrogen Receptor and Lung Cancer. doi:10.3390/ijms18081713. *Int J Mol Sci*. 2017;18(8).
- Oton AB, Belani C, Cai C, Owonikoko T, Gooding W, Ramalingam JS. Comparison of survival for non-small cell lung cancer (NSCLC) between premenopausal and postmenopausal women: An analysis of the National Surveillance, Epidemiology and End Results (SEER) Database. *J Clin Oncol*. 2006;24(18_suppl):7038. doi:10.1200/jco.2006.24.18_suppl.7038
- Slatore CG, Chien JW, Au DH, Satia JA, White E. Lung cancer and hormone replacement therapy: association in the vitamins and lifestyle study. doi:10.1200/JCO.2009.25.9739. *J Clin Oncol*. 2010;28(9):1540-6.
- Chlebowski RT, Schwartz AG, Wakelee H, et al. Oestrogen plus progestin and lung cancer in postmenopausal women (Women's Health Initiative trial): a post-hoc analysis of a randomised controlled trial. doi:10.1016/S0140-6736(09)61526-9. *Lancet*. 2009;374(9697):1243-51.
- Greiser CM, Greiser EM, Doren M. Menopausal hormone therapy and risk of lung cancer-Systematic review and meta-analysis. doi:10.1016/j.maturitas.2009.11.027. *Maturitas*. 2010;65(3):198-204.
- Smida T, Bruno TC, Stable LP. Influence of Estrogen on the NSCLC Microenvironment: A Comprehensive Picture and Clinical Implications. doi:10.3389/fonc.2020.00137. *Front Oncol*. 2020;10:137.
- Levin ER. Bidirectional signaling between the estrogen receptor and the epidermal growth factor receptor. doi:10.1210/me.2002-0368. *Mol Endocrinol*. 2003;17(3):309-17.
- Stabile LP, Lyker JS, Gubish CT, Zhang W, Grandis JR, Siegfried JM. Combined targeting of the estrogen receptor and the epidermal growth factor receptor in non-small cell lung cancer shows enhanced antiproliferative effects. -1872. *Cancer Res*. 2005;65(4):1459-1470. doi:10.1158/0008-5472.CAN-04-1872
- Fan T, Li C, He J. Prognostic value of immune-related genes and comparative analysis of immune cell infiltration in lung adenocarcinoma: sex differences. *Biol Sex Differ*. 2021;12(1):64. doi:10.1186/s13293-021-00406-y
- Wang D, Yang N, Xie S. Sex-biased ceRNA networks reveal that OSCAR can promote proliferation and migration of lung adenocarcinoma in women. *Clin Exp Pharmacol Physiol*. 2020;47(8):1350-1359. doi:10.1111/1440-1681.13318
- Danaher P, Warren S, Lu R, et al. Pan-cancer adaptive immune resistance as defined by the Tumor Inflammation Signature (TIS): results from The Cancer Genome Atlas (TCGA). *J Immunother Cancer*. 2018;6(1):63. doi:10.1186/s40425-018-0367-1
- Liang X, Li L, Fan Y. Diagnostic, Prognostic, and Immunological Roles of HELLS in Pan-Cancer: A Bioinformatics Analysis. *Front Immunol*. 2022;13:870726. doi:10.3389/fimmu.2022.870726
- Peng L, Li J, Wu J, et al. A Pan-Cancer Analysis of SMARCA4 Alterations in Human Cancers. *Front Immunol*. 2021;12:762598. doi:10.3389/fimmu.2021.762598
- Kaddoura R, Ghelani H, Alqutami F, Altaher H, Hachim M, Jan RK. Identification of Specific Biomarkers and Pathways in the Treatment Response of Infliximab for Inflammatory Bowel Disease: In-Silico Analysis. *Life (Basel)*. 2023;13(3):680. doi:10.3390/life13030680
- Sherman BT, Hao M, Qiu J, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). doi:10.1093/nar/gkac194. *Nucleic Acids Res*. 2022;50(W1):W216-21.
- Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. doi:10.1038/nprot.2008.211. *Nat Protoc*. 2009;4(1):44-57.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. doi:10.1073/pnas.0506580102. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-50.
- Remmele W, Stegner HE. [Recommendation for uniform definition of an immunoreactive score (IRS) for immunohistochemical estrogen receptor detection (ER-ICA) in breast cancer tissue]. *Pathologie*. 1987;8(3):138-140.
- Yang X, Bai Q, Chen W, et al. m^A-Dependent Modulation via IGF2BP3/MCM5/Notch Axis Promotes Partial EMT and LUAD Metastasis. *Adv Sci (Weinh)*. 2023;10(20):e2206744. doi:10.1002/adv.20206744
- Liu K, Wang L, Lou Z, et al. E2F8 exerts cancer-promoting effects by transcriptionally activating RRM2 and E2F8 knockdown synergizes with WEE1 inhibition in suppressing lung adenocarcinoma. *Biochem Pharmacol*. 2023;218:115854. doi:10.1016/j.bcp.2023.115854
- Iyer SR, Odintsov I, Schoenfeld AJ, et al. MYC Promotes Tyrosine Kinase Inhibitor Resistance in ROS1-Fusion-Positive Lung Cancer. *Mol Cancer Res*. 2022;20(5):722-734. doi:10.1158/1541-7786.MCR-22-0025
- Li C, Meng J, Zhang T. NCAPH is a prognostic biomarker and associated with immune infiltrates in lung adenocarcinoma. *Sci Rep*. 2022;12(1):9578. doi:10.1038/s41598-022-12862-6
- Ma L, Xue X, Zhang X, et al. The essential roles of m^A RNA modification to stimulate ENO1-dependent glycolysis and tumorigenesis in lung adenocarcinoma. *J Exp Clin Cancer Res*. 2022;41(1):36. doi:10.1186/s13046-021-02200-5
- Shi R, Bao X, Unger K, et al. Identification and validation of hypoxia-derived gene signatures to predict clinical outcomes and therapeutic responses in stage I lung adenocarcinoma

- patients. *Theranostics*. 2021;11(10):5061-5076. doi:10.7150/tno.56202
- Fang H, Sun Q, Zhou J, et al. m^A methylation reader IGF2BP2 activates endothelial cells to promote angiogenesis and metastasis of lung adenocarcinoma. *Mol Cancer*. 2023;22(1):99. doi:10.1186/s12943-023-01791-1
- Li G, Song Z, Wu C, et al. Downregulation of NEDD4L by EGFR signaling promotes the development of lung adenocarcinoma. *J Transl Med*. 2022;20(1):47. doi:10.1186/s12967-022-03247-4
- Huang JR, Tan GM, Li Y, Shi Z. The Emerging Role of Cables1 in Cancer and Other Diseases. doi:10.1124/mol.116.107730. *Mol Pharmacol*. 2017;92(3):240-245.
- Tan D, Kirley S, Li Q, et al. Loss of cables protein expression in human non-small cell lung cancer: a tissue microarray study. doi:10.1053/hupa.2003.26. *Hum Pathol*. 2003;34(2):143-9.
- Park DY, Sakamoto H, Kirley SD, et al. The Cables gene on chromosome 18q is silenced by promoter hypermethylation and allelic loss in human colorectal cancer. doi:10.2353/ajpath.2007.070331. *Am J Pathol*. 2007;171(5):1509-19.
- Dong Q, Kirley S, Rueda B, Zhao C, Zukerberg L, Oliva E. Loss of cables, a novel gene on chromosome 18q, in ovarian cancer. doi:10.1097/01.MP.0000084434.88269.0A. *Mod Pathol*. 2003;16(9):863-8.
- Zukerberg LR, DeBernardo RL, Kirley SD, et al. Loss of cables, a cyclin-dependent kinase regulatory protein, is associated with the development of endometrial hyperplasia and endometrial cancer. *Cancer Res*. 2004;64(1):202-8. doi:10.1158/0008-5472.CAN-03-2833
- Wu CL, Kirley SD, Xiao H, Chuang Y, Chung DC, Zukerberg LR. Cables enhances cdk2 tyrosine 15 phosphorylation by Wee1, inhibits cell growth, and is lost in many human colon and squamous cancers. *Cancer Res*. 2001;61(19):7325-7332.
- Shi Z, Li Z, Li Z, et al. Cables1 controls p21/Cip1 protein stability by antagonizing proteasome subunit alpha type 3. doi:10.1038/ncr.2014.171. *Oncogene*. 2015;34(19):2538-45.
- Tsuji K, Mizumoto K, Yamochi T, Nishimoto I, Matsuoka M. Differential effect of ik3-1/cables on p53- and p73-induced cell death. doi:10.1074/jbc.M108535200. *J Biol Chem*. 2002;277(4):2951-7.
- Wang N, Guo L, Rueda BR, Tilly JL. Cables1 protects p63 from proteasomal degradation to ensure deletion of cells after genotoxic stress. doi:10.1038/embor.2010.82. *EMBO Rep*. 2010;11(8):633-9.
- Shi Z, Park HR, Du Y, et al. Cables1 complex couples survival signaling to the cell death machinery. doi:10.1158/0008-5472.CAN-14-0036. *Cancer Res*. 2015;75(11):147-158.
- Aranson T, Pino MS, Yilmaz O, et al. Cables1 is a tumor suppressor gene that regulates intestinal tumor progression in Apc(Min) mice. doi:10.4161/cbt.25089. *Cancer Biol Ther*. 2013;14(7):672-8.
- Takeuchi K, Soda M, Togashi Y, et al. Pulmonary inflammatory myofibroblastic tumor expressing a novel fusion, PPF1BP1-ALK: reappraisal of anti-ALK immunohistochemistry as a tool for novel ALK fusion identification. *Clin Cancer Res*. 2011;17(10):3341-3348. doi:10.1158/1078-0432.CCR-11-0063
- Weishaupt H, Cančer M, Rosén G, et al. Novel cancer gene discovery using a forward genetic screen in RCAS-PDGF-driven gliomas. *Neuro-oncol*. 2023;25(1):97-107. doi:10.1093/neuonc/noac158
- Bian D, Chen Y. Bioinformatics Analysis of Prognostic Significance and Immune Characteristics of CXCL Chemokine Family in Patients with Lung Adenocarcinoma. *Computational and Mathematical Methods in Medicine*. 2022;2022:doi:10.1155/2022/3918926
- Deng J, Ma X, Ni Y, et al. Identification of CXCL5 expression as a predictive biomarker associated with response and prognosis of immunotherapy in patients with non-small cell lung cancer. *Cancer Med*. 2022;11(8):1787-1795. doi:10.1002/cam4.4567
- Zhou C-H, Ye L-P, Ye S-X, et al. Clinical significance of SOX9 in human non-small cell lung cancer progression and overall patient survival. *J Exp Clin Cancer Res*. 2012;31(1):18. doi:10.1186/1756-9966-31-18
- Jiang SS, Fang W-T, Hou Y-H, et al. Upregulation of SOX9 in lung adenocarcinoma and its involvement in the regulation of cell growth and tumorigenicity. *Clin Cancer Res*. 2010;16(17):4363-4373. doi:10.1158/1078-0432.CCR-10-0138
- De Rienzo A, Coleman MH, Yeap BY, et al. Association of RERG Expression with Female Survival Advantage in Malignant Pleural Mesothelioma. *Cancers (Basel)*. 2021;13(3):565. doi:10.3390/cancers13030565
- Liu H, Li X, Yu WQ, Liu CX. Upregulated diaphragmatic hernia rat model. *Int J Mol Med*. 2018;42(5):2373-2382. doi:10.3892/ijmm.2018.3824
- Tachibana M, Tomonoto Y, Hyakudomi R, et al. Expression and prognostic significance of EFN2 and EphB4 genes in patients with oesophageal squamous cell carcinoma. *Dig Liver Dis*. 2007;39(8):725-732. doi:10.1016/j.jld.2007.05.013
- Tang Z, Xie H, Heier C, et al. Enhanced monoacylglycerol lipolysis by ABHD6 promotes NSCLC pathogenesis. *EBioMedicine*. 2020;53:102696. doi:10.1016/j.ebiom.2020.102696
- Yu DH, Huang JY, Liu XP, et al. Effects of hub genes on the clinicopathological and prognostic features of lung adenocarcinoma. *Oncol Lett*. 2020;19(2):1203-1214.
- She K, Yu S, He S, Wang W, Chen B. CircRNA 0009043 suppresses non-small-cell lung cancer development via targeting the miR-148a-3p/DNAJB4 axis. *Biomark Res*. 2022;10(1):61. doi:10.1186/s40364-022-00407-y
- Tsai M-F, Wang C-C, Chang G-C, et al. A new tumor suppressor DnaJ-like heat shock protein, HLJ1, and survival of patients with non-small-cell lung carcinoma. *J Natl Cancer Inst*. 2006;98(12):825-838. doi:10.1093/jnci/djj229
- Peralta-Arrieta I, Trejo-Villegas OA, Armas-López L, et al. Failure to EGFR-TKI-based therapy and tumoural progression are promoted by MEOX2/GLI1-mediated epigenetic regulation of EGFR in the human lung cancer. *Eur J Cancer*. 2022;160:189-205. doi:10.1016/j.ejca.2021.10.032
- Cortese R, Hartmann O, Berlin K, Eckhardt F. Correlative gene expression and DNA methylation profiling in lung development nominate new biomarkers in lung cancer. *Int J Biochem Cell Biol*. 2008;40(8):1494-1508. doi:10.1016/j.biocel.2007.11.018
- Ávila-Moreno F, Armas-López L, Álvarez-Moran AM, et al. Overexpression of MEOX2 and TWIST1 is associated with H3K27me3 levels and determines lung cancer chemoresistance and prognosis. *PLoS One*. 2014;9(12):e114104. doi:10.1371/journal.pone.0114104
- Nagelberg AL. *Integrative genomic analyses of lung adenocarcinoma from never-smokers*. University of British Columbia; 2022.
- Yin J, Ma R, Wang T, Su Z, Xu H. The predominant Th17 phenotype in gastric cancers may be associated with HLX expression decrease blocking Th1 cell specific T-box transcription factor. 2012;
- Zhu XY, Guo QY, Zhu M, et al. HLX affects cell cycle and proliferation in AML cells via the JAK/STAT signaling pathway. *Oncol Lett*. 2020;20(2):1888-1896. doi:10.3892/ol.2020.11718
- Chiu H-W, Chang J-S, Lin H-Y, et al. FBXL7 Upregulation Predicts a Poor Prognosis and Associates with a Possible Mechanism for Paclitaxel Resistance in Ovarian Cancer. *J Clin Med*. 2018;7(10):330. doi:10.3390/jcm7100330